2017

Boundary crossing probabilities for the cumulative sample mean

Singham, Dashi I.

D.I. Singham & M.P. Atkinson (2017). Boundary Crossing Probabilities for the Cumulative Sample Mean. Probability in the Engineering and Informational Sciences, 2017. doi:10.1017/S026996481700002X

Probability in the Engineering and Informational Sciences, page 1 of 21, 2017.
doi:10.1017/S026996481700002X

BOUNDARY CROSSING PROBABILITIES FOR THE CUMULATIVE SAMPLE MEAN

Dashi I. Singham
Operations Research Department,
Naval Postgraduate School,
Monterey, CA 93943,
USA
E-mail: dsingham@nps.edu

Michael P. Atkinson
Operations Research Department,
Naval Postgraduate School,
Monterey, CA 93943,
USA
E-mail: mpatkins@nps.edu

We develop a new measure of reliability for the mean behavior of a process by calculating the probability the cumulative sample mean will stay within a given distance from the true mean over a period of time. This probability is derived using boundary-crossing properties of Brownian bridges. We derive finite sample results for independent and identically distributed normal data, limiting results for data meeting a functional central limit theorem, and draw parallels to standard normal confidence intervals. We deliver numerical results for i.i.d., dependent, and queueing processes.

Keywords: applied probability, reliability theory, simulation, brownian bridge

1. INTRODUCTION

We calculate the probability the sample mean of a time series stays within some fixed distance from its long-term mean over a given period of time. We derive separate results for two variants of the "long-term mean": the sample mean calculated after m observations have been collected, $\overline{Y}_m$, and the "true" process mean μ. That is, given a time series Y_i for $i = 1, 2, \ldots$, we define two expressions:

$$P^m := P\left(\bigcap_{k\le j\le m}\left\{\left|\frac{1}{m}\sum_{i=1}^{m}Y_i - \frac{1}{j}\sum_{i=1}^{j}Y_i\right| \le \delta\right\}\right), \quad P := P\left(\bigcap_{j\ge k}\left\{\left|\mu - \frac{1}{j}\sum_{i=1}^{j}Y_i\right| \le \delta\right\}\right), \tag{1.1}$$

for some $1 \le k < m$. The parameter m denotes the number of samples used to calculate the long-term mean, and the first expression of (1.1), P^m, is the probability the cumulative sample mean between samples k and m stays within distance δ from $\overline{Y}_m$. The second expression, P, represents the probability that the sample mean stays within distance δ from its true mean μ after an initial sample size k. We also remove the absolute value signs from (1.1) and derive the probability the sample mean stays within a given distance from the true mean in one direction.

These expressions P^m and P have parallels to confidence intervals, which deliver an interval with a specified coverage probability for the true mean using one fixed sample size. P is the probability an interval centered at the true mean will contain the cumulative sample mean path for an infinite sample size. In practice, stability of the cumulative sample mean over time is a more conservative risk metric than a confidence interval based on a fixed sample size. P and P^m capture the probability of short term deviations of the sample mean from the true mean and help determine sample sizes needed for the cumulative mean to stabilize according to some desired precision δ. Being able to deliver consistent performance close to the true mean is often just as important as having a good long-term mean, which may or may not be realized in the short-term. Certainly in many settings such as healthcare and defense, poor short-term performance is not acceptable even if long-term average performance is good. A conservative practitioner should ensure a high value of P in addition to a high value of μ estimated by a confidence interval.

Confidence intervals can calibrate the calculation of P by providing a good estimate of μ. A stochastic process will not constantly exhibit mean performance μ due to random variation, but there may be some range of mean performance $\mu \pm \delta$ that can be managed. For example, suppose a factory manufactures W units of a product each day to serve a random demand of Y_i units per day, where Y_i is independent and identically distributed (i.i.d.) with mean μ and variance σ^2. If demand is smaller than W on a given day, excess inventory can be held to meet demand at a later date. Shortages are unacceptable and the manager wants to control the probability that this occurs. Suppose the system starts on day one with no inventory. If on any day i, the realized cumulative mean demand $\overline{Y}_i$ is greater than W, demand has not been met on at least one of the days. Given that μ can be estimated and $\mu < W$, then in the long run the factory can meet demand *on average*, but not necessarily every day. Using $\delta = W - \mu$, P estimates the probability a shortage will occur after the kth day. This information can be used to determine if the manufacturer should invest in better technology to boost production, or supplement the initial inventory to cover daily variation in early days. Other applications include whether a reservoir serving a community will run out of water, or the possibility a firm will need to pay overtime given current staffing levels and work requirements.

In calculating P and P^m, we assume the underlying time series $\{Y_i,\ i \geq 1\}$ meets the conditions for a functional central limit theorem (FCLT). For i.i.d. normal data, we derive exact results for all possible values of k and m. For FCLT data we derive conditions to obtain limiting probabilities as $k, m \to \infty$, and display numerical results for dependent and non-normal data to demonstrate this convergence in the limit. We evaluate P and P^m by rewriting them as functions of a standardized time series, which under some conditions converges to a Brownian bridge. Modifying the standardization function enables adaptation of boundary-crossing results for Brownian bridges to derive a closed-form lower bound for (1.1). The lower bound occurs from discretization error, and this error can be made arbitrarily small with larger sample sizes.

We summarize the contributions of the paper. Section 2 derives lower bounds for P^m and P when Y_i is i.i.d. normal. We demonstrate a relationship between P and the standard normal confidence interval coverage probability $1 - \alpha$ and find the probability the cumulative sample mean stays within δ of the true mean is approximately $1 - 2\alpha$. Section 3 derives limiting expressions for (1.1) for FCLT data, and includes numerical analysis for different stochastic processes to illustrate convergence. Section 4 concludes.

2. CALCULATING P^m AND P

In this section, we derive lower bounds for P and P^m when the underlying data are i.i.d. normal, and compare the results to confidence interval coverage probabilities.

2.1. Preliminaries

We rely on the standardization technique introduced by Schruben [15] primarily used for developing conference interval estimators for the mean, μ. Schruben constructed standardized time series and proved that they converge to a Brownian bridge using Billingsley [5]. Many researchers have used standardized time series methods to estimate the mean and variance of data that have some degree of serial dependency (Batur, Goldsman, and Kim [4], Goldsman, Meketon, and Schruben [9]). We take a different approach and exploit the properties of Brownian bridges to derive the probability the cumulative mean stays within some distance from the long-term mean and the true mean. When deriving properties of P and P^m, we assume a known variance σ^2. If δ is expressed as $\delta'\sigma$ for some $\delta' > 0$, then σ will cancel from the final expression. We acknowledge that obtaining a consistent estimate of the variance can be difficult for dependent data. Much effort has been devoted to obtaining variance estimates of dependent data, and we refer the reader to Alexopoulos [2] for a review.

We assume the data, $Y_1, \ldots, Y_m$, is stationary with $E[Y_i] = \mu$ and $Var[Y_i] = \sigma_Y^2$. Define the sample mean for each $k = 1, \ldots, m$ as $\overline{Y}_k = (1/k)\sum_{i=1}^{k} Y_i$. Let σ^2 be the asymptotic variance constant defined in Glynn and Iglehart [8] as $\lim_{k\to\infty} k\,Var[\overline{Y}_k]$. Assume $0 < \sigma^2 < \infty$, and denote R_j as the autocovariance of Y_i and Y_{i+j} for all $i > 0$. The following expresses the relationship between these terms:

$$\sigma^2 = \sigma_Y^2 + 2\sum_{j=1}^{\infty} R_j. \tag{2.1}$$

Schruben [15] defines a standardized time series as

$$X_m(t) = \frac{\lfloor mt \rfloor \left((1/m)\sum_{i=1}^{m} Y_i - (1/\lfloor mt \rfloor)\sum_{i=1}^{\lfloor mt \rfloor} Y_i \right)}{\sigma\sqrt{m}}, \quad t \in [0,1]. \tag{2.2}$$

We require the following FCLT assumption.

ASSUMPTION 2.1: *(FCLT): Define*

$$C(t,m) = \frac{mt\mu - \sum_{i=1}^{\lfloor mt \rfloor} Y_i}{\sigma\sqrt{m}}, \quad t \geq 0$$

as a function in Skorohod space $D[0,\infty]$, which consists of right continuous functions with left limits. Then, $C(t,m)$ converges weakly to standard Brownian motion as $m \to \infty$. See Section 4.4 of Whitt [18] for details.

Schruben [15] proves that under the assumptions of a FCLT, $X_m(t)$ converges weakly to $B(t)$ as $m \to \infty$, where $B(t)$ is a standard Brownian bridge. Examples of data that meet the assumptions of a FCLT include strictly stationary ϕ-mixing data, where observations relatively far apart are independent (see Billingsley [5]) and strictly stationary strong mixing data (see Glynn and Iglehart [8]).

While Assumption 2.1 is mild, producing exact results for P^m using finite values of k and m requires that the joint distribution of the points $X_m(j/m), j = 1, \ldots, m$ have the same joint distribution as the points $B(j/m)$ of a standard Brownian bridge. This occurs when the data are i.i.d. normal (see Singham and Atkinson [16] for a proof). The next subsection derives lower bounds for P^m and P when the data are i.i.d. normal. Section 3 derives the limiting probabilities when the data satisfies Assumption 2.1.

2.2. Lower Bound Derivation

In this section, we derive lower bound expressions for P and P^m. Figure 1 illustrates cumulative mean behavior relative to δ. Using (2.2) in conjunction with the i.i.d. normal data assumption, we rewrite the left expression in (1.1) in terms of a standardized time series $X_m(t)$ and a Brownian bridge $B(t)$:

$$P^m = P\left(\bigcap_{k\le j\le m}\left\{\left|\sigma X_m(\frac{j}{m})\frac{\sqrt{m}}{j}\right| \le \delta\right\}\right) = P\left(\bigcap_{k\le j\le m}\left\{\left|\sigma X_m(\frac{j}{m})\right| \le \delta\frac{j}{\sqrt{m}}\right\}\right) \quad \textbf{(2.3)}$$

$$= P\left(\bigcap_{k\le j\le m}\left\{\left|\sigma B(\frac{j}{m})\right| \le \delta\frac{j}{\sqrt{m}}\right\}\right) \quad \textbf{(2.4)}$$

$$\ge P\left(\bigcap_{t\in[(k/m),1]}\left\{|\sigma B(t)| \le \delta\sqrt{m}t\right\}\right) \equiv P_L^m. \quad \textbf{(2.5)}$$

With i.i.d. normal data, (2.3) to (2.4) holds because the joint distribution of the points $X_m(j/m), j = k, \ldots, m$ has the same distribution as the points $B(j/m)$ of a standard Brownian bridge.

To move from (2.4) to (2.5), note that the events in (2.4) are a subset of the events in (2.5) as (2.4) evaluates the crossing condition for discrete points only. We denote the lower bound in (2.5) P_L^m, which is a conservative estimate of P^m when evaluating process risk. It is important to note that we produce a lower bound rather than an equality only because

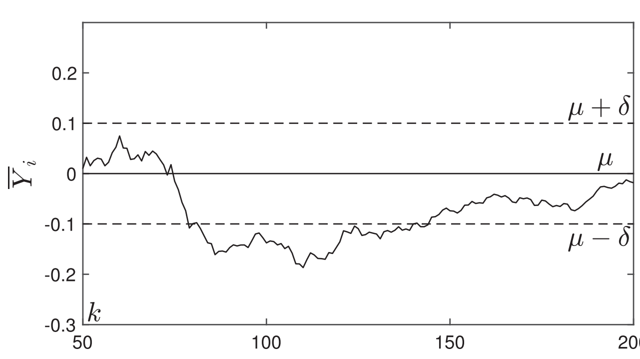

FIGURE 1. Given an initial sample size k, P is the probability that the cumulative sample mean stays within distance δ from the true mean μ (the bounds are the dashed lines).

of the discretization error as we move from a discrete process in (2.4) to a continuous one in (2.5). The value of the lower bound in (2.5) is calculated analytically, and is easier to obtain than estimating P^m directly. Section 3 presents the limiting results for P^m as $k, m \to \infty$, $\delta \to 0$, and discusses the order of the discretization error.

The calculation of (2.5) builds on previous work that studies linear boundary-crossing probabilities for Brownian bridges such as Abundo [1] and Scheike [14]. Abundo [1] derives the probability a Brownian bridge ever leaves two symmetric linear bounds that have non-zero intercepts at $t = 0$. In our case, the slope of these linear bounds is $\pm\delta\sqrt{m}$. Whereas the intercept at $t = 0$ is zero, we start the process at $t = k/m$, which yields a non-zero intercept. In practice, an experiment would require some initial k samples to calculate an estimate of the sample mean. For a general reference on boundary-crossing problems for Brownian motion, see Karatzas and Shreve [10].

To calculate (2.5), condition on the value of $\sigma B(k/m)$ and make this the starting point of a new Brownian bridge. Moving forward from time k/m, the process can be viewed as a new Brownian bridge with initial position x and final position zero. Let $B_x^{k/m}(t)$ be a Brownian bridge over $[0, 1 - k/m]$ that starts at x/σ and ends at zero. The unconditional starting position of $\sigma B(k/m)$ is normally distributed with mean 0 and variance $\sigma^2 \frac{k}{m}(1 - \frac{k}{m})$. The bounds for the cumulative mean in (2.5) have non-zero intercepts $\pm\delta k/\sqrt{m}$ at $t = k/m$, and the time range for the Brownian bridge is now $[0, 1 - k/m]$. We compute P_L^m by integrating over the possible initial position x at time k/m:

$$P_L^m = \int_{x=-\delta k/\sqrt{m}}^{\delta k/\sqrt{m}} P\left(\bigcap_{t\in[0,1-(k/m)]} \{|\sigma B_x^{k/m}(t)| \le \delta k/\sqrt{m} + \delta\sqrt{m}t\}\right) \times N\left(x, 0, \sigma^2 \frac{k}{m}\left(1 - \frac{k}{m}\right)\right) dx, \quad \textbf{(2.6)}$$

where $N(x, \mu, \sigma^2)$ is the normal probability density function with mean μ and variance σ^2 evaluated at x. We write P_L^m as $P_L^m(\delta, \sigma, k, m)$ to denote dependence on the problem parameters. For brevity, all proofs appear in the appendix. The first theorem gives the direct evaluation of (2.6). Let Φ be the cumulative distribution function of the standard normal distribution.

THEOREM 2.2: *Under the assumption of i.i.d. normal data, the probability that the sample mean stays within distance δ from its long-term mean $\overline{Y}_m$ over the range $j = k, \ldots, m$ has a lower bound*

$$P_L^m(\delta, \sigma, k, m) \le P\left(\bigcap_{k\le j\le m} \left\{\left|\frac{1}{m}\sum_{i=1}^m Y_i - \frac{1}{j}\sum_{i=1}^j Y_i\right| \le \delta\right\}\right)$$

where

$$P_L^m(\delta, \sigma, k, m) = 1 - 4\sum_{i=1}^\infty \left(\Phi\left(\frac{\delta\sqrt{k}}{\sigma\sqrt{1 - \frac{k}{m}}}(4i - 1)\right) - \Phi\left(\frac{\delta\sqrt{k}}{\sigma\sqrt{1 - \frac{k}{m}}}(4i - 3)\right)\right). \quad \textbf{(2.7)}$$

The proof of Theorem 2.2 appears in Appendix A. In the next theorem we state the parallel lower bound P_L for P and use the notation $P_L(\delta, \sigma, k)$ to remove the dependence on m. As expected, $P_L^m \to P_L$ as $m \to \infty$.

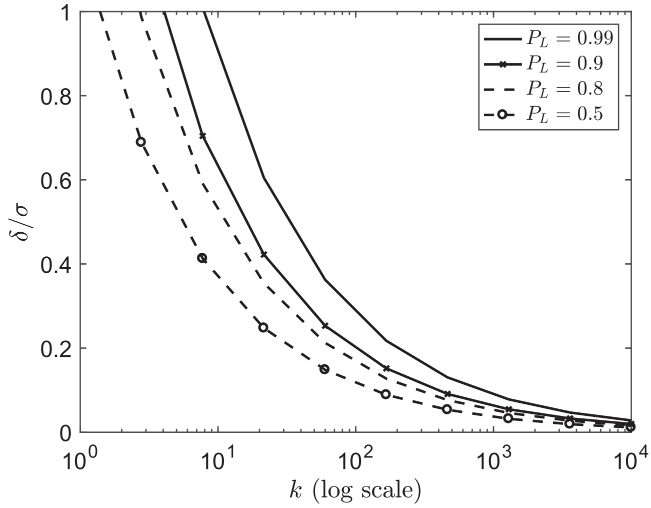

FIGURE **2.** Calculation of (2.8) showing P_L for a range of values of k and δ/σ.

THEOREM 2.3: *Under the assumption of i.i.d. normal data, the probability that the sample mean stays within distance δ from its true mean μ for all $j \geq k$ has a lower bound*

$$P_L(\delta, \sigma, k) \leq P\left(\bigcap_{j \geq k} \left\{\left|\mu - \frac{1}{j}\sum_{i=1}^{j} Y_i\right| \leq \delta\right\}\right)$$

where

$$P_L(\delta, \sigma, k) = 1 - 4\sum_{i=1}^{\infty}\left(\Phi\left(\frac{\delta\sqrt{k}}{\sigma}(4i-1)\right) - \Phi\left(\frac{\delta\sqrt{k}}{\sigma}(4i-3)\right)\right). \qquad \textbf{(2.8)}$$

The proof of Theorem 2.3 appears in Appendix A. As in Theorem 2.2, the lower bound results only from the discretization error.

Like confidence intervals, important tradeoffs for P exist between δ, σ and k. Figure 2 plots several contours of P_L, calculated using (2.8) for a variety of combinations of k and δ/σ. Because δ and σ only appear as δ/σ in (2.8) we condense them to one term, and δ could be defined as $\delta'\sigma$ for $\delta' > 0$ to avoid estimating the variance directly. On the x-axis, as k increases, P_L increases because the sample mean is less likely to deviate from the true mean at larger sample sizes. On the y-axis, P_L increases with the ratio δ/σ because we have a larger bound relative to the variance. As σ increases, this ratio decreases and P_L also decreases. Inspection of (2.8) reveals that $\frac{\delta\sqrt{k}}{\sigma}$ remains constant for a given contour, and thus one must scale δ proportional to $1/\sqrt{k}$ to maintain a desired value of P_L.

2.3. Parallels to confidence intervals

Next, we present a simple, analytically tractable approximation for P_L that has a negligible error for realistic scenarios and allows for parallels to confidence intervals.

COROLLARY 2.4: *The values of P_L^m and P_L in Theorems 2.2 and 2.3 have lower bounds*

$$4\Phi\left(\frac{\delta\sqrt{k}}{\sigma\sqrt{1-\frac{k}{m}}}\right)-3 \le P_L^m(\delta,\sigma,k,m), \quad \text{and} \quad 4\Phi\left(\frac{\delta\sqrt{k}}{\sigma}\right)-3 \le P_L(\delta,\sigma,k). \tag{2.9}$$

The upper bound on the difference between P_L^m, P_L and these lower bounds are

$$4\left(1-\Phi\left(3\frac{\delta\sqrt{k}}{\sigma\sqrt{1-\frac{k}{m}}}\right)\right), \quad \text{and} \quad 4\left(1-\Phi\left(3\frac{\delta\sqrt{k}}{\sigma}\right)\right) \tag{2.10}$$

respectively.

The proof of Corollary 2.4 appears in Appendix B. The lower bound gap in (2.10) is extremely small for large values of P_L. If $P_L(\delta,\sigma,k)=0.3$, the gap is within 0.011; if $P_L(\delta,\sigma,k)=0.5$, the gap is less than 0.0011; if $P_L(\delta,\sigma,k)=0.8$, the gap is less than 2×10^{-6}. In realistic applications, P_L should be close to 1, and thus for all practical purposes these lower bounds can be used in place of P_L.

We next consider the one-sided boundary to compute the probability that $\overline{Y}_j$ will ever be less than its long-term mean, or greater than its long-term mean, with allowed deviation δ in one direction. That is, the sample mean should stay within $[\overline{Y}_m-\delta,\infty]$ or $[-\infty,\overline{Y}_m+\delta]$. In the limiting case, these bounds are $[\mu-\delta,\infty]$ or $[-\infty,\mu+\delta]$. We denote the one-sided lower bounds P_L' to distinguish them from the two-sided bounds P_L. Under the assumption of i.i.d. normal data, the probability that the sample mean stays within distance δ (on one side) from its long-term mean $\overline{Y}_m$ over the range $j=k,\ldots,m$ has a lower bound

$$P_L'(\delta,\sigma,k,m)=2\Phi\left(\frac{\delta\sqrt{k}}{\sigma\sqrt{1-\frac{k}{m}}}\right)-1 \le P\left(\bigcap_{k\le j\le m}\left\{\frac{1}{m}\sum_{i=1}^{m}Y_i-\frac{1}{j}\sum_{i=1}^{j}Y_i\le\delta\right\}\right).$$

The probability that the sample mean stays within distance δ (on one side) from μ for all $j\ge k$ has a lower bound

$$P_L'(\delta,\sigma,k)=2\Phi\left(\frac{\delta\sqrt{k}}{\sigma}\right)-1 \le P\left(\bigcap_{j\ge k}\left\{\mu-\frac{1}{j}\sum_{i=1}^{j}Y_i\le\delta\right\}\right). \tag{2.11}$$

The proof deriving P_L' appears in Singham and Atkinson [16]. We highlight two interesting relationships related to the one-sided expression in (2.11). First, $P_L'(\delta,\sigma,k)$ equals the two-sided confidence interval coverage probability for a sample size of k

$$CI(\delta,\sigma,k) \;\equiv\; P\left(\left|\mu-\frac{1}{k}\sum_{i=1}^{k}Y_i\right|\le\delta\right).$$

Second, the relationship between the one-sided result in (2.11) and the two-sided lower bound in Corollary 2.4 yields:

$$P_L(\delta,\sigma,k)\approx 2P_L'(\delta,\sigma,k)-1.$$

This is the same functional relationship between one-sided and two-sided confidence intervals. Furthermore, since $P_L'(\delta,\sigma,k)=CI(\delta,\sigma,k)$ we can connect a confidence interval to P using the relationship $P_L(\delta,\sigma,k) \approx 2CI(\delta,\sigma,k) - 1$. Given a sample size k, variance σ^2, and precision δ that produces a $1-\alpha$ confidence interval, the same parameters would generate a P_L probability of $1-2\alpha$.

When the variance of the data is known, a fixed-sample size that will deliver a confidence interval with desired coverage $1-\alpha$ and precision smaller than δ can be chosen using the formula

$$k \geq \left(\frac{\Phi^{-1}(1-(\alpha/2))\sigma}{\delta}\right)^2, \tag{2.12}$$

where Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution. In a similar fashion, P_L can be used to estimate the number of samples needed to obtain a probabilistic guarantee on the cumulative sample mean staying within a given distance from μ. In particular Corollary 2.4 produces a similar expression to (2.12) for choosing a sample size k to obtain a probability that the sample mean will continue to stay within δ of μ:

$$k \geq \left(\frac{\Phi^{-1}(1-(\alpha/4))\sigma}{\delta}\right)^2. \tag{2.13}$$

Computing the sample size to obtain a P_L guarantee is equivalent to computing the sample size required for a confidence interval coverage probability guarantee using $\alpha/2$. Because the two-sided measure is more strict than a standard confidence interval, condition (2.13) requires a larger starting sample size than condition (2.12). In the one-sided case, the value of k (using P'_L from (2.11)) is the same as the value needed to satisfy (2.12).

3. LIMITING RESULTS

In this section, we derive a limiting value of P^m and P for FCLT data. For dependent and non-normal data with finite sample sizes, the joint distributions of $X_m(j/m)$ and $B(j/m), j=1,\ldots,m$ will not be the same as moving from (2.3) to (2.4) requires. We describe how k and δ must behave as $m\to\infty$ to obtain a limiting value of P^m. Examination of (2.3)–(2.5) reveals that if the empirical process crosses the boundary, it will likely do so close to time k/m because the slope of the boundary scales with $\sqrt{m}$. Thus we need $X_m(t)$ to approximate a Brownian bridge reasonably well over small time scales (i.e., close to sample k). This implies it is not merely enough to have large m, we must also have large k. To derive an interesting limiting result with both $m\to\infty$ and $k\to\infty$, we also send $\delta\to 0$. Otherwise, for fixed δ, P_L^m and P^m will both approach 1 as $k,m\to\infty$. In the next theorem we give the conditions for a limiting result to hold.

Theorem 3.1: *Let Assumption 2.1 hold. For $k(m)=sm$ for some $s\in(0,1)$ and $\delta(m)=\frac{\theta}{\sqrt{m}}$ for some $\theta>0$,*

$$\lim_{m\to\infty} P\left(\bigcap_{k(m)\leq j\leq m}\left\{\left|\frac{1}{m}\sum_{i=1}^{m}Y_i - \frac{1}{j}\sum_{i=1}^{j}Y_i\right| \leq \delta(m)\right\}\right) = P_L\left(\theta,\sigma,\frac{s}{1-s}\right). \tag{3.1}$$

The proof of Theorem 3.1 appears in Appendix C. For finite m, Theorem 3.1 allows us to approximate P^m for data meeting a FCLT. The inputs δ, k, and m enable the calculation of the auxiliary variables $s=k/m$ and $\theta=\delta\sqrt{m}$. Using these auxiliary variables implies $P_L(\theta,\sigma,(s/1-s)) = P_L^m(\delta,\sigma,k,m)$ for any m. Thus the practical implication of

Theorem 3.1 is that $P_L^m(\delta, \sigma, k, m)$ should closely approximate P^m for dependent and/or non-normal data that satisfy the FCLT, *provided* that m and k are large enough. We next provide the analogous result for P.

THEOREM 3.2: *Let Assumption 2.1 hold. For $\delta(k) = (\alpha/\sqrt{k})$ for some $\alpha > 0$,*

$$\lim_{k\to\infty} P\left(\bigcap_{j\geq k}\left\{\left|\mu - \frac{1}{j}\sum_{i=1}^{j} Y_i\right| \leq \delta(k)\right\}\right) = P_L(\alpha, \sigma, 1). \tag{3.2}$$

The proof of Theorem 3.2 appears in Appendix C. The practical implication of Theorem 3.2 is that $P_L(\delta, \sigma, k)$ approximates P for FCLT data, provided k is large enough.

We briefly discuss of the effect of discretization error on the difference between P_L^m (P_L) and P^m (P) for finite sample sizes. When choosing $\delta = \theta/\sqrt{m}$ in Theorem 3.1, the boundaries $\delta\sqrt{m}t$ become fixed at θt. Prior research exists on the order of the error between the probability a discrete skeleton of a Brownian motion crosses linear bounds and the probability a continuous Brownian motion crosses the same bounds. This error decreases at rate $O(m^{-1/2})$ (Nagaev [13], Lerche and Siegmund [12], Fu and Wu [7]). Consequently when δ, k, and m change at the appropriate rates, the error associated with using P_L^m instead of P^m is $O(m^{-1/2})$ (or $O(k^{-1/2})$ as k must increase as well).

The rest of the section applies Theorem 3.1 to show the convergence for different data types. We compute P^m numerically by simulating Y_i and determining whether the cumulative mean crosses the boundary. The results reveal the effects of discretization error, non-normality, dependence and sample size.

The i.i.d. normal results isolate the effect of the lower bound discretization error by removing any distributional or finite-sample issues. Figure 3 shows plots of P_L^m and P^m for m=100,000 and different values of δ. Each plot displays four different calculations across different values of k. The solid line is P_L^m as calculated according to (2.7). The remaining dashed lines show P^m calculated using normal, lognormal and Pareto distributions. Tests with exponential and uniform data yielded P^m values very close to those for normal data.

For small values of k, P^m is close to zero; as k increases, the values of P^m increase and approach one. The lower bound gap due to discretization error is demonstrated in the difference between P_L^m and P^m for normally distributed data, and is noticeable for small k. As we increase k by a factor of 10 and decrease δ by a factor of $\sqrt{10}$ in each plot, we observe

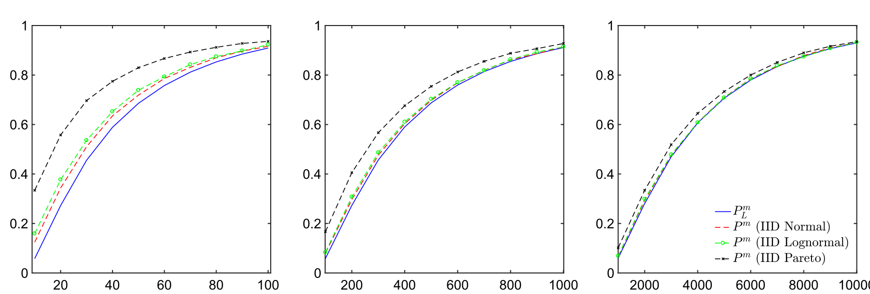

FIGURE **3.** Values of P_L^m and P^m for independent data, with $m = 100{,}000$. We have $\delta = 0.20\sigma$ (left), $\delta = 0.0632\sigma$ (center), and $\delta = 0.02\sigma$ (right). The P^m lines are from simulated data, all with $\sigma = 1$: $N(0, 1)$, $Lognormal(\hat{\mu} = 0, \hat{\sigma} = 0.693)$ and $Pareto(scale = 1, shape = 2.84)$.

convergence to the limiting result as suggested by Theorem 3.1. The Pareto distribution is highly non-normal and hence takes much longer to converge than the other distributions.

Next, we consider highly-dependent processes. Figure 4 displays P_L^m for AR(1) data with ϵ_i as i.i.d. normal, $m = 1{,}000{,}000$ and several values of ρ and k. As suggested by Theorem 3.1, we decrease δ as we increase k by keeping the product $\delta\sqrt{k}$ constant as we move from left to right. We display results for high auto-correlation coefficients $\rho \in \{\pm 0.5, \pm 0.9\}$ to demonstrate worst-case conditions. In addition to discretization issues, the dependence leads to P_L^m being a poor approximation for P^m for small k. Furthermore, for negative values of ρ it is possible that P^m is actually less than P_L^m. P_L^m is only guaranteed to be a lower bound of P for i.i.d. normal data. For small values of k and significant dependence, a Brownian bridge does not approximate the standardized time series well. Computing simple confidence intervals for data of this type is also difficult, and we refer the reader to Tafazzoli and Wilson [17] for confidence interval estimation methods for difficult data types. The main point of this analysis is that while crossing probabilities can be calculated in the limit when the standardized time series has true Brownian properties, it becomes more difficult in smaller sample sizes without making distributional assumptions.

We conclude with a similar analysis for the waiting times in an M/M/1 queue. Customers arrive according to a Poisson process with rate λ and service times are i.i.d. exponential random variables with mean $1/\mu$. The traffic intensity $\rho = \frac{\lambda}{\mu} \in (0,1)$ drives the dependence structure of the data. Figure 5 displays the results for m=10,000,000 in a

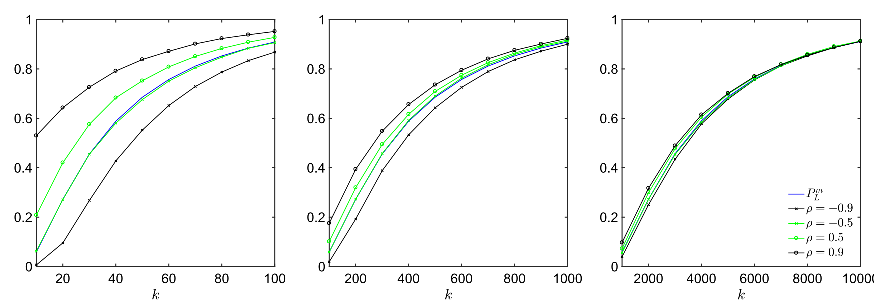

FIGURE 4. Values of P_L^m and P^m for AR(1) data with $\rho = \pm 0.5, 0.9$ and $m = 1{,}000{,}000$. We have $\delta = 0.20\sigma$ (left), $\delta = 0.0632\sigma$ (center), and $\delta = 0.02\sigma$ (right). The parameter k increases by a factor of 10 for each figure moving left to right.

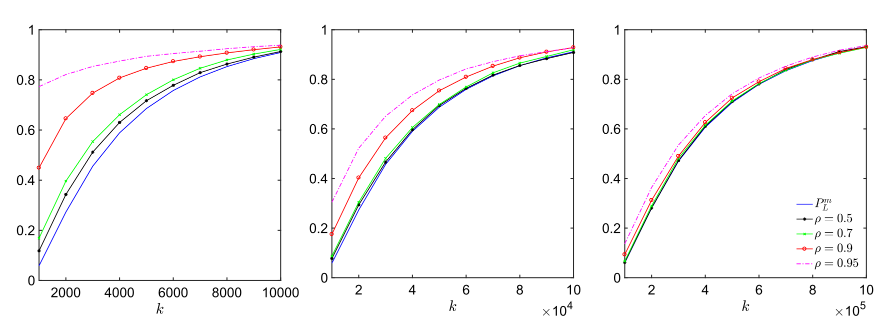

FIGURE 5. Values of P_L and P for the waiting times of an M/M/1 queue with $\rho = 0.5, 0.7, 0.9, 0.95$ and $m = 10{,}000{,}000$. We have $\delta = 0.02\sigma$ (left), $\delta = 0.00632\sigma$ (center), and $\delta = 0.002\sigma$ (right). The parameter k increases by a factor of 10 for each figure moving left to right.

similar format to Figure 4. For $k \in [10^3, 10^4]$, P_L^m is a reasonable fit for $\rho = 0.5, 0.7$, although a nontrivial gap exists between the two curves and the P_L^m curve. The difference between P_L^m and P^m is very large for $\rho = 0.9, 0.95$. As we increase k and decrease δ, the P curves converge to the P_L^m curve. For $k \in [10^5, 10^6]$, the curves for $\rho = 0.5, 0.7, 0.9$ nearly coincide with the P_L^m curve, and the $\rho = 0.95$ curve has a relatively minor deviation. Given the highly correlated nature of waiting times, it is not surprising that we need a very large value of k as ρ approaches 1 for P_L^m to be a good approximation for P^m. The results of this section suggest that P_L^m may be a lower bound for P^m for positively correlated data, which could prove useful for smaller sample sizes.

4. CONCLUSION

We introduce a measure of reliability for mean process behavior by deriving the probability that the cumulative sample mean stays within a given distance from its long-term mean over a period of time beginning at time k, where the long-term mean can be $\overline{Y}_m$ or μ. We compute a lower bound on this probability, where the lower bound occurs because of discretization error. If the data is i.i.d. normal, then the lower bound applies for finite sample sizes. As k and m approach infinity, this lower bound becomes exact for data meeting a FCLT.

The probability is a function of the allowable distance from the true mean, the variance of the underlying data, and the initial number of samples taken. We quantify the trade-offs between these different parameters and draw parallels to confidence interval coverage probabilities. This measure enables evaluation of the relative stability of sample mean performance by assessing process behavior over a range of time, rather than just at a snapshot in time.

References

1. Abundo, M. (2002). Some conditional crossing results of Brownian motion over a piecewise-linear boundary. *Statistics & Probability Letters* 58(2): 131–145.
2. Alexopoulos, C. (2006). A comprehensive review of methods for simulation output analysis. In L.P. Perrone, F.P. Wieland, J. Liu, B.G. Lawson, D.M. Nicol, & R.M. Fujimoto (eds.), *Proceedings of the 2006 Winter Simulation Conference*. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., pp. 168–178.
3. Atkinson, M.P. & Singham, D.I. (2015).Multidimensional hitting time results for Brownian bridges with moving hyperplanar boundaries. *Statistics and Probability Letters* 100: 85–92.
4. Batur, D., Goldsman, D., & Kim, S.-H. (2009). An improved standardized time series Durbin–Watson variance estimator for steady-state simulation. *Operations Research Letters* 37(4): 285–289.
5. Billingsley, P. (1968). *Convergence of probability measures*. New York: Wiley.
6. Doob, J.L. (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *The Annals of Mathematical Statistics* 20(3): 393–403.
7. Fu, J.C. & Wu, T.-L. (2010). Linear and nonlinear boundary crossing probabilities for Brownian motion and related processes. *Journal of Applied Probability* 47(4): 1058–1071.
8. Glynn, P.W. & Iglehart, D.L. (1990). Simulation output analysis using standardized time series. *Mathematics of Operations Research* 15(1): 1–16.
9. Goldsman, D., Meketon, M., & Schruben, L. (1990). Properties of standardized time series weighted area variance estimators. *Management Science* 36(5): 602–612.
10. Karatzas, I. & Shreve, S. (1998). *Brownian motion and stochastic calculus*. New York, NY: Springer.
11. Klenke, A. (2013). *Probability theory: a comprehensive course*. London: Springer Science & Business Media.
12. Lerche, H.R. & Siegmund, D. (1987). Approximate Exit Probabilities for a Brownian Bridge on a Short Time Interval, and Applications. Technical report, Stanford University.

13. Nagaev, S.V. (1970). On the speed of convergence in a boundary problem. I. *Theory of Probability & Its Applications* 15(2): 163–186.
14. Scheike, T.H. (1992). A boundary-crossing result for Brownian motion. *Journal of Applied Probability* 29(2): 448–453.
15. Schruben, L. (1983). Confidence interval estimation using standardized time series. *Operations Research* 31(6): 1090–1108.
16. Singham, D.I. & Atkinson, M.P. (2015). Construction of Cumulative Mean Bounds for Simulation Output Analysis. Technical Report NPS-OR-15-004, Naval Postgraduate School.
17. Tafazzoli, A. & Wilson, J.R. (2010). Skart: a skewness-and autoregression-adjusted batch-means procedure for simulation analysis. *IIE Transactions* 43(2): 110–128.
18. Whitt, W. (2002). *Stochastic-process limits: an introduction to stochastic-process limits and their application to queues*. New York, NY: Springer Science & Business Media.

APPENDIX A: PROOF OF THEOREM 2.2 AND 2.3

The lower bounds P_L^m and P_L for i.i.d. normal data are derived in the following sections.

Proof of Theorem 2.2: First, rewrite P_L^m in (2.5) through conditioning on the location of $B(t)$ at time $t = k/m$ as done in (2.6):

$$\begin{aligned} P_L^m(\delta,\sigma,k,m) &= P\left(\bigcap_{t\in[\frac{k}{m},1]} \{|\sigma B(t)| \le \delta\sqrt{m}t\}\right) \\ &= \int_x P\left(\bigcap_{t\in[(k/m),1]} \{|\sigma B(t)| \le \delta\sqrt{m}t | \sigma B(k/m) = x\}\right) N\left(x, 0, \sigma^2\frac{k}{m}\left(1-\frac{k}{m}\right)\right) dx \\ &= \int_{-\frac{\delta k}{\sqrt{m}}}^{\frac{\delta k}{\sqrt{m}}} P\left(\bigcap_{t\in[0,1-(k/m)]} \left\{|\sigma B_x^{k/m}(t)| \le \frac{\delta k}{\sqrt{m}} + \delta\sqrt{m}t\right\}\right) \\ &\quad \times N\left(x, 0, \sigma^2\frac{k}{m}\left(1-\frac{k}{m}\right)\right) dx. \end{aligned} \tag{A.1}$$

The remainder of this section simplifies (A.1) to

$$P_L^m(\delta,\sigma,k,m) = 1 - 4\sum_{j=1}^{\infty}\left(\Phi\left(\frac{\delta\sqrt{k}}{\sigma\sqrt{1-(k/m)}}(4j-1)\right) - \Phi\left(\frac{\delta\sqrt{k}}{\sigma\sqrt{1-(k/m)}}(4j-3)\right)\right). \tag{A.2}$$

We rely on a result from Doob [6] that computes the probability a one-dimensional standard Brownian motion stays between two lines for all time. This result defines a function $G(\cdot,\cdot,\cdot,\cdot)$ as

$$G(\alpha,\beta,\gamma,\lambda) = 1 - \sum_{j=1}^{\infty}(e^{-2A_j} + e^{-2B_j} - e^{-2C_j} - e^{-2D_j}), \tag{A.3}$$

with

$$\begin{aligned} A_j &= j^2\gamma\lambda + (j-1)^2\alpha\beta + j(j-1)(\gamma\beta+\lambda\alpha), \\ B_j &= (j-1)^2\gamma\lambda + j^2\alpha\beta + j(j-1)(\gamma\beta+\lambda\alpha), \\ C_j &= j^2(\gamma\lambda+\alpha\beta) + j(j-1)\gamma\beta + j(j+1)\lambda\alpha, \\ D_j &= j^2(\gamma\lambda+\alpha\beta) + j(j+1)\gamma\beta + j(j-1)\lambda\alpha. \end{aligned} \tag{A.4}$$

Doob's result also appears as Theorem 3.1 in Abundo [1]. For $\alpha, \beta, \gamma, \lambda > 0$, Doob [6] proves

$$P\left(\bigcap_{t\geq 0}\{-(\alpha t+\beta)\leq W(t)\leq \gamma t+\lambda\}\right) = G(\alpha,\beta,\gamma,\lambda), \tag{A.5}$$

where $W(t)$ is a standard Brownian motion. Atkinson and Singham [3] derive a similar relationship for the probability that a Brownian bridge that starts at location $-\beta < x_0 < \lambda$ at time 0, ends at location $-(\alpha T+\beta) < x_T < (\gamma T+\lambda)$ at time T, and having variance parameter σ^2 stays between two asymmetric linear boundaries, $-(\alpha t+\beta)$ and $\gamma t+\lambda$. If we denote $\tilde{B}(t)$ as the Brownian bridge with the properties described in the previous sentence, then we have the following probability of interest

$$P\left(\bigcap_{t\in[0,T]}\{-(\alpha t+\beta)\leq \tilde{B}(t)\leq \gamma t+\lambda\}\right) = G\left(\frac{\lambda - x_0}{\sigma}, \frac{\lambda+\gamma T - x_T}{\sigma T}, \frac{x_0+\beta}{\sigma}, \frac{x_T+\beta+\alpha T}{\sigma T}\right).$$

The symmetric case appears in Abundo [1]. We use this result to substitute the following into (A.1)

$$\begin{aligned}
&P\left(\bigcap_{t\in[0,1-(k/m)]}\left\{|\sigma B_x^{k/m}(t)|\leq \frac{\delta k}{\sqrt{m}}+\delta\sqrt{m}t\right\}\right)\\
&\quad = G\left(\frac{(\delta k/\sqrt{m})-x}{\sigma}, \frac{\delta\sqrt{m}}{\sigma(1-(k/m))}, \frac{x+(\delta k/\sqrt{m})}{\sigma}, \frac{\delta\sqrt{m}}{\sigma(1-(k/m))}\right).
\end{aligned}$$

Therefore, we can now rewrite (A.1) as

$$\begin{aligned}
P_L(\delta,\sigma,k,m) = \int_{-(\delta k/\sqrt{m})}^{(\delta k/\sqrt{m})} & G\left(\frac{(\delta k/\sqrt{m})-x}{\sigma}, \frac{\delta\sqrt{m}}{\sigma(1-(k/m))}, \frac{x+(\delta k/\sqrt{m})}{\sigma}, \frac{\delta\sqrt{m}}{\sigma(1-(k/m))}\right)\\
&\times N\left(x, 0, \sigma^2\frac{k}{m}\left(1-\frac{k}{m}\right)\right)dx.
\end{aligned}$$

Next, we change variables via the relationship $x = (\delta k/\sqrt{m})u$ where u varies in $[-1, 1]$. This not only changes the limits of integration, but it changes the variance of the normal density we are integrating against so that

$$\begin{aligned}
&P_L(\delta,\sigma,k,m)\\
&\quad = \int_{-1}^{1} G\left(\frac{((\delta k/\sqrt{m})-(\delta k/\sqrt{m})u)}{\sigma}, \frac{\delta\sqrt{m}}{\sigma(1-(k/m))}, \frac{((\delta k/\sqrt{m})u+(\delta k/\sqrt{m}))}{\sigma}, \frac{\delta\sqrt{m}}{\sigma(1-(k/m))}\right)\\
&\qquad \times N\left(\frac{\delta k}{\sqrt{m}}u, 0, \sigma^2\frac{k}{m}\left(1-\frac{k}{m}\right)\right)\frac{\delta k}{\sqrt{m}}du\\
&\quad = \int_{-1}^{1} G\left(\frac{(\delta k/\sqrt{m})(1-u)}{\sigma}, \frac{\delta\sqrt{m}}{\sigma(1-(k/m))}, \frac{(\delta k/\sqrt{m})(1+u)}{\sigma}, \frac{\delta\sqrt{m}}{\sigma(1-(k/m))}\right)\\
&\qquad \times N\left(u, 0, \frac{\sigma^2}{\delta^2 k}\left(1-\frac{k}{m}\right)\right)du.
\end{aligned} \tag{A.6}$$

We next focus on the G function. Note that the second and fourth arguments of $G(\cdot,\cdot,\cdot,\cdot)$ in (A.6) are the same. By inspection of (A.3) and (A.4), we see that the G function consists of products of two out of the four input arguments. Two of the six possible combinations do not appear: the first and third arguments and the second and fourth arguments. These characteristics of the G function and its inputs allow us to rewrite (A.6) as

$$\int_{-1}^{1} G\left(\frac{\delta k(1-u)}{\sigma}, \frac{\delta}{\sigma(1-(k/m))}, \frac{\delta k(1+u)}{\sigma}, \frac{\delta}{\sigma(1-(k/m))}\right) N\left(u, 0, \frac{\sigma^2}{\delta^2 k}\left(1-\frac{k}{m}\right)\right) du. \tag{A.7}$$

For the function $G(\alpha, \beta, \gamma, \lambda)$ in (A.7) we have

$$\begin{aligned}
\gamma\lambda = \gamma\beta &= \frac{\delta^2 k(1+u)}{\sigma^2(1-(k/m))} = q(1+u), \\
\alpha\beta = \lambda\alpha &= \frac{\delta^2 k(1-u)}{\sigma^2(1-(k/m))} = q(1-u),
\end{aligned}$$

where $q = \frac{\delta^2 k}{\sigma^2(1-(k/m))}$. Next, substitute these values into (A.4) to obtain

$$\begin{aligned}
A_j &= j^2 q(1+u) + (j-1)^2 q(1-u) + j(j-1)q(1+u) + j(j-1)q(1-u) \\
&= j(2j-1)q(1+u) + (j-1)(2j-1)q(1-u) \\
&= (2j-1)\left(jq(1+u) + jq(1-u) - q(1-u)\right) \\
&= j(2j-1)\left(q(1+u) + q(1-u)\right) - (2j-1)q(1-u) \\
&= j(2j-1)2q - (2j-1)q(1-u),
\end{aligned}$$

$$\begin{aligned}
B_j &= (j-1)^2 q(1+u) + j^2 q(1-u) + j(j-1)q(1+u) + j(j-1)q(1-u) \\
&= (j-1)(2j-1)q(1+u) + j(2j-1)j^2 q(1-u) \\
&= j(2j-1)2q - (2j-1)q(1+u),
\end{aligned}$$

$$\begin{aligned}
C_j &= j^2 q(1+u) + j^2 q(1-u) + j(j-1)q(1+u) + j(j+1)q(1-u) \\
&= j(2j-1)q(1+u) + j(2j+1)q(1-u) \\
&= 2j^2\left(q(1+u) + q(1-u)\right) - jq(1+u) + jq(1-u) \\
&= 2j^2(2q) - 2jqu \\
&= j(2j-1)2q + 2jq - 2jqu \\
&= j(2j-1)2q + 2jq(1-u),
\end{aligned}$$

$$\begin{aligned}
D_j &= j^2 q(1+u) + j^2 q(1-u) + j(j+1)q(1+u) + j(j-1)q(1-u) \\
&= j(2j+1)q(1+u) + j(2j-1)q(1-u) \\
&= 2j^2(2q) + 2jqu \\
&= j(2j-1)2q + 2jq(1+u).
\end{aligned}$$

Writing out the $G(\cdot,\cdot,\cdot,\cdot)$ function explicitly from (A.7) gives

$$P_L(\delta, \sigma, k, m) = \int_{-1}^{1}\left(1 - \sum_{j=1}^{\infty}\left(e^{-2A_j} + e^{-2B_j} - e^{-2C_j} - e^{-2D_j}\right)\right) N\left(u, 0, q^{-1}\right) du. \tag{A.8}$$

By comparing expressions for A_j, B_j, C_j and D_j, we have

$$\int_{-1}^{1} \sum_{j=1}^{\infty} e^{-2A_j} N\left(u, 0, q^{-1}\right) du = \int_{-1}^{1} \sum_{j=1}^{\infty} e^{-2B_j} N\left(u, 0, q^{-1}\right) du, \tag{A.9}$$

$$\int_{-1}^{1} \sum_{j=1}^{\infty} e^{-2C_j} N\left(u, 0, q^{-1}\right) du = \int_{-1}^{1} \sum_{j=1}^{\infty} e^{-2D_j} N\left(u, 0, q^{-1}\right) du. \tag{A.10}$$

To prove these relationships, first interchange the sum and integration by the Fubini–Tonelli theorem and write one term in the sum in (A.9) as

$$\int_{-1}^{1} e^{-2A_j} N(u, 0, q^{-1}) du = \int_{-1}^{1} \exp\left(-2\left(j(2j-1)2q - (2j-1)q(1-u)\right)\right) \frac{\exp\left(-(u^2/2q^{-1})\right)}{\sqrt{2\pi q^{-1}}} du.$$

If we define $z = -u$ then we have

$$-\int_{1}^{-1} \exp\left(-2\left(j(2j-1)2q - (2j-1)q(1+z)\right)\right) \frac{\exp\left(-(z^2/2q^{-1})\right)}{\sqrt{2\pi q^{-1}}} dz = \int_{-1}^{1} e^{-2B_j} N(z, 0, q^{-1}) dz.$$

We can make the same argument with C_j and D_j to show (A.10). We can thus re-write (A.8) in terms of only A_j and C_j so that

$$\begin{aligned}
P_L(\delta, \sigma, k, m) &= \int_{-1}^{1} \left(1 - 2\sum_{j=1}^{\infty}(e^{-2A_j} - e^{-2C_j})\right) N\left(u, 0, q^{-1}\right) du \\
&= \int_{-1}^{1} N\left(u, 0, q^{-1}\right) du - \sum_{j=1}^{\infty} \int_{-1}^{1} \left(2e^{-2A_j} - 2e^{-2C_j}\right) N\left(u, 0, q^{-1}\right) du \\
&= \left(2\Phi\left(\sqrt{q}\right) - 1\right) - 2\sum_{j=1}^{\infty} \int_{-1}^{1} \left(e^{-2A_j} - e^{-2C_j}\right) \frac{\exp\left(-(u^2/2q^{-1})\right)}{\sqrt{2\pi q^{-1}}} du. \qquad \textbf{(A.11)}
\end{aligned}$$

First, we examine the terms in (A.11) containing A_j:

$$\begin{aligned}
&\int_{-1}^{1} e^{-2A_j} \frac{\exp\left(-(u^2/2q^{-1})\right)}{\sqrt{2\pi q^{-1}}} du \\
&= \int_{-1}^{1} \exp\left(-2\left(j(2j-1)2q - (2j-1)q(1-u)\right)\right) \frac{\exp\left(-(u^2/2q^{-1})\right)}{\sqrt{2\pi q^{-1}}} du \\
&= \exp\left(-2(2j-1)^2 q\right) \int_{-1}^{1} \exp\left(-2(2j-1)qu\right) \frac{\exp\left(-(u^2/2q^{-1})\right)}{\sqrt{2\pi q^{-1}}} du \\
&= \exp\left(-2(2j-1)^2 q\right) \int_{-1}^{1} \exp\left(-(2j-1)\frac{4u}{2q^{-1}}\right) \frac{\exp\left(-(u^2/2q^{-1})\right)}{\sqrt{2\pi q^{-1}}} du.
\end{aligned}$$

Combining exponential terms in the integral yields

$$
\begin{aligned}
&= \exp\left(-2\left((2j-1)^2 q\right)\right)\int_{-1}^{1}\frac{\exp\left(-(u^2+(2j-1)4u)/2q^{-1}\right)}{\sqrt{2\pi q^{-1}}}du\\
&= \exp\left(-2\left((2j-1)^2 q\right)\right)\int_{-1}^{1}\frac{\exp\left(-((u+2(2j-1))^2-4(2j-1)^2)/2q^{-1}\right)}{\sqrt{2\pi q^{-1}}}du\\
&= \exp\left(-2\left((2j-1)^2 q\right)\right)\int_{-1}^{1}\frac{\exp\left(-((u+2(2j-1))^2)/2q^{-1}\right)\exp\left(4(2j-1)^2/2q^{-1}\right)}{\sqrt{2\pi q^{-1}}}du.
\end{aligned}
$$

The first and last exponential terms cancel, which produces the final expression for the A_j term:

$$
\begin{aligned}
\int_{-1}^{1}\frac{\exp\left(-((u+2(2j-1))^2)/2q^{-1}\right)}{\sqrt{2\pi q^{-1}}}du &= \Phi\left(\sqrt{q}(1+2(2j-1))\right)-\Phi\left(\sqrt{q}(-1+2(2j-1))\right)\\
&= \Phi\left(\sqrt{q}(4j-1)\right)-\Phi\left(\sqrt{q}(4j-3)\right).
\end{aligned}
$$

Going through similar steps for the C_j term of of (A.11) yields

$$
\begin{aligned}
\int_{-1}^{1} e^{-2C_j}\frac{\exp\left(-(u^2/2q^{-1})\right)}{\sqrt{2\pi q^{-1}}}du &= \int_{-1}^{1}\exp\left(-2\left(j(2j-1)2q+2jq(1-u)\right)\right)\frac{\exp\left(-(u^2/2q^{-1})\right)}{\sqrt{2\pi q^{-1}}}du\\
&= \exp\left(-2\ (2j)^2 q\right)\int_{-1}^{1}\exp\left((2j)\frac{4u}{2q^{-1}}\right)\frac{\exp\left(-(u^2/2q^{-1})\right)}{\sqrt{2\pi q^{-1}}}du.
\end{aligned}
$$

Combining exponential terms in the integral again yields:

$$
\begin{aligned}
&= \exp\left(-2\left((2j)^2 q\right)\right)\int_{-1}^{1}\frac{\exp\left(-(u^2-(2j)4u)/2q^{-1}\right)}{\sqrt{2\pi q^{-1}}}du\\
&= \exp\left(-2\left((2j)^2 q\right)\right)\int_{-1}^{1}\frac{\exp\left(-((u-2(2j))^2-4(2j)^2)/2q^{-1}\right)}{\sqrt{2\pi q^{-1}}}du\\
&= \exp\left(-2\left((2j)^2 q\right)\right)\int_{-1}^{1}\frac{\exp\left(-((u-2(2j))^2/2q^{-1})\right)\exp\left((4(2j)^2/2q^{-1})\right)}{\sqrt{2\pi q^{-1}}}du\\
&= \int_{-1}^{1}\frac{\exp\left(-((u-2(2j))^2/2q^{-1})\right)}{\sqrt{2\pi q^{-1}}}du = \Phi\left(\sqrt{q}(1-2(2j))\right)-\Phi\left(\sqrt{q}(-1-2(2j))\right)\\
&= \Phi\left(\sqrt{q}(1-4j)\right)-\Phi\left(\sqrt{q}(-1-4j)\right).
\end{aligned}
$$

Combining everything into (A.11) gives $P_L^m(\delta,\sigma,k,m) =$

$$
\begin{aligned}
&(2\Phi\left(\sqrt{q}\right)-1)-2\sum_{j=1}^{\infty}\left(\Phi\left(\sqrt{q}(4j-1)\right)-\Phi\left(\sqrt{q}(4j-3)\right)\right)\\
&\quad+2\sum_{j=1}^{\infty}\left(\Phi\left(\sqrt{q}(1-4j)\right)-\Phi\left(\sqrt{q}(-1-4j)\right)\right)
\end{aligned}
$$

$$= \left(2\Phi\left(\sqrt{q}\right)-1\right) - 2\sum_{j=1}^{\infty}\left(\Phi\left(\sqrt{q}(4j-1)\right)-\Phi\left(\sqrt{q}(4j-3)\right)\right)$$

$$- 2\sum_{j=1}^{\infty}\left(\Phi\left(\sqrt{q}(4j-1)\right)-\Phi\left(\sqrt{q}(4j+1)\right)\right)$$

$$= \left(2\Phi\left(\sqrt{q}\right)-1\right) - 2\sum_{j=1}^{\infty}\left(\Phi\left(\sqrt{q}(4j-1)\right)-\Phi\left(\sqrt{q}(4j-3)\right)\right)$$

$$- 2\sum_{j=1}^{\infty}\left(\Phi\left(\sqrt{q}(4j-1)\right)-\Phi\left(\sqrt{q}(4j-3)\right)\right) - 2\sum_{j=1}^{\infty}\left(\Phi\left(\sqrt{q}(4j-3)\right)-\Phi\left(\sqrt{q}(4j+1)\right)\right)$$

$$= \left(2\Phi\left(\sqrt{q}\right)-1\right) - 4\sum_{j=1}^{\infty}\left(\Phi\left(\sqrt{q}(4j-1)\right)-\Phi\left(\sqrt{q}(4j-3)\right)\right)$$

$$+ 2\sum_{j=1}^{\infty}\left(\Phi\left(\sqrt{q}(4j+1)\right)-\Phi\left(\sqrt{q}(4j-3)\right)\right)$$

$$= \left(2\Phi\left(\sqrt{q}\right)-1\right) - 4\sum_{j=1}^{\infty}\left(\Phi\left(\sqrt{q}(4j-1)\right)-\Phi\left(\sqrt{q}(4j-3)\right)\right)$$

$$+ 2\lim_{L\to\infty}\left(\Phi\left(\sqrt{q}(4L+1)\right)-\Phi\left(\sqrt{q}(4\cdot 1-3)\right)\right)$$

$$= \left(2\Phi\left(\sqrt{q}\right)-1\right) - 4\sum_{j=1}^{\infty}\left(\Phi\left(\sqrt{q}(4j-1)\right)-\Phi\left(\sqrt{q}(4j-3)\right)\right) + 2\left(1-\Phi\left(\sqrt{q}\right)\right).$$

Canceling terms and substituting for q produces the final answer

$$P_L(\delta,\sigma,k,m) = 1 - 4\sum_{j=1}^{\infty}\left(\Phi\left(\frac{\delta\sqrt{k}}{\sigma\sqrt{1-(k/m)}}(4j-1)\right) - \Phi\left(\frac{\delta\sqrt{k}}{\sigma\sqrt{1-(k/m)}}(4j-3)\right)\right). \quad \blacksquare$$

Proof of Theorem 2.3: We proceed in a similar fashion as the proof of Theorem 2.2. We first define a similar empirical process to $X_m(t)$ in (2.2):

$$Z_k(t) = \frac{\lfloor kt\rfloor\left(\mu - (1/\lfloor kt\rfloor)\sum_{i=1}^{\lfloor kt\rfloor} Y_i\right)}{\sigma\sqrt{k}}, \quad t \geq 0. \tag{A.12}$$

By Donsker's Theorem (e.g., Section 4.4 of Whitt [18]) $Z_k(\cdot)$ converges to a standard Brownian motion. Taking steps similar to (2.3)–(2.5) yields

$$P := P\left(\bigcap_{j\geq k}\left\{\left|\mu - \frac{1}{j}\sum_{i=1}^{j} Y_i\right| \leq \delta\right\}\right) = P\left(\bigcap_{j\geq k}\left\{\left|\sigma Z_k(\tfrac{j}{k})\frac{\sqrt{k}}{j}\right| \leq \delta\right\}\right) \tag{A.13}$$

$$= P\left(\bigcap_{j\geq k}\left\{\left|\sigma Z_k(\tfrac{j}{k})\right| \leq \delta\frac{j}{\sqrt{k}}\right\}\right) \tag{A.14}$$

$$= \lim_{N\to\infty} P\left(\bigcap_{k\leq j\leq N}\left\{\left|\sigma Z_k(\tfrac{j}{k})\right| \leq \delta\frac{j}{\sqrt{k}}\right\}\right) \tag{A.15}$$

$$= \lim_{N\to\infty} P\left(\bigcap_{k\le j\le N}\left\{\left|\sigma W(\frac{j}{k})\right| \le \delta\frac{j}{\sqrt{k}}\right\}\right) \tag{A.16}$$

$$\ge \lim_{N\to\infty} P\left(\bigcap_{t\in[1,\frac{N}{k}]}\left\{|\sigma W(t)| \le \delta\sqrt{k}t\right\}\right) \tag{A.17}$$

$$= P\left(\bigcap_{t\ge 1}\left\{|\sigma W(t)| \le \delta\sqrt{k}t\right\}\right) = P_L(\delta,\sigma,k). \tag{A.18}$$

(A.15) follows from the dominated convergence theorem. The i.i.d. normality assumption allows us to move from (A.15) to (A.16). (A.17) is a lower bound for (A.16) because of the discretization bias. Finally (A.18) follows from the dominated convergence theorem. We now compute $P_L(\delta,\sigma,k)$ relying on parallel results from the proof of Theorem 2.2 where necessary:

$$P_L(\delta,\sigma,k) = \int_x P\left(\bigcap_{t\ge 1}\{|\sigma W(t)| \le \delta\sqrt{k}t \mid \sigma W(1) = x\}\right) N\left(x,0,\sigma^2\right) dx$$

$$= \int_{-\delta\sqrt{k}}^{\delta\sqrt{k}} P\left(\bigcap_{t\ge 0}\left\{|\sigma W_x(t)| \le \delta\sqrt{k} + \delta\sqrt{k}t\right\}\right) N\left(x,0,\sigma^2\right) dx \tag{A.19}$$

where $W_x(t)$ denotes a standard Brownian motion that starts at x/σ at time 0. The unconditional position at $W(1)$ is normally distributed with mean 0 and variance σ^2. We rewrite the first term in the integral of (A.19)

$$P\left(\bigcap_{t\ge 0}\left\{|\sigma W_x(t)| \le \delta\sqrt{k} + \delta\sqrt{k}t\right\}\right)$$

$$= P\left(\bigcap_{t\ge 0}\left\{-\left(\frac{\delta\sqrt{k}+x}{\sigma} + \frac{\delta\sqrt{k}}{\sigma}t\right) \le W(t) \le \frac{\delta\sqrt{k}-x}{\sigma} + \frac{\delta\sqrt{k}}{\sigma}t\right\}\right). \tag{A.20}$$

The right-hand side of (A.20) has the same form as (A.5). Consequently

$$P\left(\bigcap_{t\ge 0}\left\{|\sigma W_x(t)| \le \delta\sqrt{k} + \delta\sqrt{k}t\right\}\right) = G\left(\frac{\delta\sqrt{k}}{\sigma}, \frac{\delta\sqrt{k}+x}{\sigma}, \frac{\delta\sqrt{k}}{\sigma}, \frac{\delta\sqrt{k}-x}{\sigma}\right), \tag{A.21}$$

where $G(\cdot,\cdot,\cdot,\cdot)$ is defined in (A.3)–(A.4). Substituting (A.21) into (A.19) and proceeding in the same manner as in the proof of Theorem 2.2 yields

$$P_L(\delta,\sigma,k) = \int_{-1}^{1} G\left(\frac{\delta}{\sigma}, \frac{\delta k(1+u)}{\sigma}, \frac{\delta}{\sigma}, \frac{\delta k(1-u)}{\sigma}\right) N\left(u,0,\frac{\sigma^2}{\delta^2 k}\right) du. \tag{A.22}$$

Notice the strong similarities between (A.22) and (A.7). The identical steps from the proof of Theorem 2.2 after equation (A.7) yield the desired result:

$$P_L(\delta,\sigma,k) = 1 - 4\sum_{i=1}^{\infty}\left(\Phi\left(\frac{\delta\sqrt{k}}{\sigma}(4i-1)\right) - \Phi\left(\frac{\delta\sqrt{k}}{\sigma}(4i-3)\right)\right). \qquad \blacksquare$$

APPENDIX B: PROOF OF COROLLARY 2.4

Since the proof is the same for P_L^m and P_L, we focus on P_L. Re-write P_L as

$$P_L = 1 - 4\sum_{i=1}^{\infty} \left(\Phi\left(a(4i-1)\right) - \Phi\left(a(4i-3)\right)\right), \tag{B.23}$$

where $a = (\delta\sqrt{k}/\sigma)$. We will show that

$$4\Phi(a) - 3 \leq P_L,$$

and bound the difference $P_L - (4\Phi(a) - 3)$. First, pull the $i = 1$ term out from (B.23):

$$P_L = 1 - 4\Phi\left(3a\right) + 4\Phi\left(a\right) - 4\sum_{i=2}^{\infty} \left(\Phi\left(a(4i-1)\right) - \Phi\left(a(4i-3)\right)\right).$$

Next, rearrange terms and add and subtract 4 to obtain

$$P_L = 4\Phi\left(a\right) - 4 + 1 - 4\Phi\left(3a\right) + 4 - 4\sum_{i=2}^{\infty} \left(\Phi\left(a(4i-1)\right) - \Phi\left(a(4i-3)\right)\right).$$

Split the expression into three terms

$$\begin{aligned} P_L = {} & 4\Phi\left(a\right) - 3 \\ & + 4(1 - \Phi\left(3a\right)) \\ & - 4\sum_{i=2}^{\infty} \left(\Phi\left(a(4i-1)\right) - \Phi\left(a(4i-3)\right)\right). \end{aligned} \tag{B.24}$$

To show that $4\Phi\left(a\right) - 3$ is a lower bound, it suffices to show the sum of the second and third lines is positive. First rewrite $1 - \Phi\left(3a\right)$ as

$$\begin{aligned} 1 - \Phi\left(3a\right) &= \sum_{i=2}^{\infty}(\Phi\left((2i+1)a\right) - \Phi\left((2i-1)a\right)) \\ &= \sum_{i=2,even}^{\infty} (\Phi\left((2i+1)a\right) - \Phi\left((2i-1)a\right)) + \sum_{i=3,odd}^{\infty} (\Phi\left((2i+1)a\right) - \Phi\left((2i-1)a\right)) \\ &= \sum_{i=2}^{\infty}(\Phi\left((4i-3)a\right) - \Phi\left((4i-5)a\right)) + \sum_{i=2}^{\infty}(\Phi\left((4i-1)a\right) - \Phi\left((4i-3)a\right)). \end{aligned} \tag{B.25}$$

The second summation in (B.25) appears in the third line of (B.24). Substituting (B.25) into (B.24) yields

$$\begin{aligned} P_L &= 4\Phi\left(a\right) - 3 \\ &\quad + 4\left(\sum_{i=2}^{\infty}(\Phi\left((4i-3)a\right) - \Phi\left((4i-5)a\right)) + \sum_{i=2}^{\infty}(\Phi\left((4i-1)a\right) - \Phi\left((4i-3)a\right))\right) \\ &\quad - 4\sum_{i=2}^{\infty} \left(\Phi\left(a(4i-1)\right) - \Phi\left(a(4i-3)\right)\right) \\ &= 4\Phi\left(a\right) - 3 + 4\left(\sum_{i=2}^{\infty}(\Phi\left((4i-3)a\right) - \Phi\left((4i-5)a\right))\right), \end{aligned}$$

which delivers the result.

APPENDIX C: PROOF OF THEOREMS 3.1 AND 3.2

The proofs for Theorems 3.1 and 3.2 are similar.

PROOF OF THEOREM 3.1: The following Lemma is needed.

LEMMA 3.1: *For any $\sigma, \theta > 0$ and $s \in (0,1)$*

$$\lim_{m\to\infty} P\left(\bigcap_{t\in[s,1]} \{|\sigma X_m(t)| \le \theta t\}\right) = P\left(\bigcap_{t\in[s,1]} \{|\sigma B(t)| \le \theta t\}\right). \qquad \textbf{(C.26)}$$

PROOF: First note by examination of (2.5) and (2.7) that we can write

$$P\left(\bigcap_{t\in[s,1]} \{|\sigma B(t)| \le \theta t\}\right) = P_L^m(\theta, \sigma, s, 1). \qquad \textbf{(C.27)}$$

We will next show that the strictness of the inequality does not impact our result, so

$$P\left(\bigcap_{t\in[s,1]} \{|\sigma B(t)| \le \theta t\}\right) = P\left(\bigcap_{t\in[s,1]} \{|\sigma B(t)| < \theta t\}\right). \qquad \textbf{(C.28)}$$

To do this we note that $P_L^m(\theta, \sigma, s, 1)$ is continuous in θ by inspection of (2.7). This follows because we can interchange limits and sums in (2.7) as the normal c.d.f. terms can be bounded by a decaying exponential function of i. Using (C.27), create a sandwich inequality

$$P_L^m(\theta - \epsilon, \sigma, s, 1) \le P\left(\bigcap_{t\in[s,1]} \{|\sigma B(t)| < \theta t\}\right) \le P_L^m(\theta + \epsilon, \sigma, s, 1), \qquad \textbf{(C.29)}$$

for any $\epsilon > 0$. Taking the limit of (C.29) as $\epsilon \to 0$ and leveraging the continuity of $P_L^m(\theta, \sigma, s, 1)$ in θ yields (C.28).

The final step uses the Portmanteau theorem for equivalency of weak convergence (see, e.g., Chapter 13 of Klenke [11]). Because $X_m(t)$ converges weakly to $B(t)$, Portmanteau provides the following two relationships

$$\limsup_{m\to\infty} P\left(\bigcap_{t\in[s,1]} \{|\sigma X_m(t)| \le \theta t\}\right) \le P\left(\bigcap_{t\in[s,1]} \{|\sigma B(t)| \le \theta t\}\right) \qquad \textbf{(C.30)}$$

$$\liminf_{m\to\infty} P\left(\bigcap_{t\in[s,1]} \{|\sigma X_m(t)| < \theta t\}\right) \ge P\left(\bigcap_{t\in[s,1]} \{|\sigma B(t)| < \theta t\}\right). \qquad \textbf{(C.31)}$$

By condition (C.28), conditions (C.30)–(C.31) form another sandwich inequality, which produces the desired result

$$\lim_{m\to\infty} P\left(\bigcap_{t\in[s,1]} \{|\sigma X_m(t)| \le \theta t\}\right) = P\left(\bigcap_{t\in[s,1]} \{|\sigma B(t)| \le \theta t\}\right). \qquad \textbf{(C.32)}$$

■

To prove Theorem 3.1 set $k(m) = sm$, $\delta(m) = \frac{\theta}{\sqrt{m}}$, and manipulate the left-hand side of (3.1):

$$P\left(\bigcap_{k(m)\le j\le m}\left\{\left|\sigma X_m(\frac{j}{m})\right| \le \delta(m)\frac{j}{\sqrt{m}}\right\}\right) = P\left(\bigcap_{t\in[\frac{k(m)}{m},1]}\{|\sigma X_m(t)| \le \delta(m)\sqrt{m}t\}\right) \quad \textbf{(C.33)}$$

$$= P\left(\bigcap_{t\in[s,1]}\{|\sigma X_m(t)| \le \theta t\}\right). \quad \textbf{(C.34)}$$

By Lemma 3.1 replace $X_m(t)$ in (C.34) with $B(t)$ in the limit:

$$\lim_{m\to\infty} P\left(\bigcap_{k(m)\le j\le m}\left\{\left|\sigma X_m(\frac{j}{m})\right| \le \delta(m)\frac{j}{\sqrt{m}}\right\}\right) = \lim_{m\to\infty} P\left(\bigcap_{t\in[s,1]}\{|\sigma X_m(t)| \le \theta t\}\right)$$

$$= P\left(\bigcap_{t\in[s,1]}\{|\sigma B(t)| \le \theta t\}\right). \quad \textbf{(C.35)}$$

Combining (C.35) with (C.27) yields

$$\lim_{m\to\infty} P\left(\bigcap_{k(m)\le j\le m}\left\{\left|\frac{1}{m}\sum_{i=1}^{m} Y_i - \frac{1}{j}\sum_{i=1}^{j} Y_i\right| \le \delta(m)\right\}\right)$$

$$= \lim_{m\to\infty} P\left(\bigcap_{k(m)\le j\le m}\left\{\left|\sigma X_m(\frac{j}{m})\right| \le \delta(m)\frac{j}{\sqrt{m}}\right\}\right) = P_L^m(\theta,\sigma,s,1). \quad \textbf{(C.36)}$$

The final piece follows by inspection of (2.7) and (2.8), which reveals that $P_L^m(\theta,\sigma,s,1) = P_L(\theta,\sigma,(s/1-s))$. Substituting this into (C.36) completes the proof. ■

Proof of Theorem 3.2: We omit the details that are similar to the previous proofs. Following the proof of Theorem 2.3 in Appendix A, use the empirical process $Z_k(t)$ defined in (A.12) to obtain

$$P\left(\bigcap_{j\ge k}\left\{\left|\mu - \frac{1}{j}\sum_{i=1}^{j} Y_i\right| \le \delta(k)\right\}\right) = \left(\bigcap_{j\ge k}\left\{\left|\sigma Z_k(\frac{j}{k})\right| \le \delta(k)\frac{j}{\sqrt{k}}\right\}\right)$$

$$= \left(\bigcap_{t\ge 1}\left\{|\sigma Z_k(t)| \le \delta(k)\sqrt{k}t\right\}\right)$$

$$= \left(\bigcap_{t\ge 1}\{|\sigma Z_k(t)| \le \alpha t\}\right).$$

There exists a relationship parallel to (C.26) from Lemma 3.1:

$$\lim_{k\to\infty} P\left(\bigcap_{t\ge 1}\{|\sigma Z_k(t)| \le \alpha t\}\right) = P\left(\bigcap_{t\ge 1}\{|\sigma W(t)| \le \alpha t\}\right) = P_L(\alpha,\sigma,1),$$

where the last equality follows by inspection of (A.18). The intermediary steps follow from the proof of Theorem 3.1. ■